E82-10119

SR-LO-00483
JSC-16384 160886

CR

AgRISTARS

A Joint Program for Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing

Supporting Research

September 1980

PROBABILISTIC CLUSTER LABELING OF IMAGERY DATA

C. B. Chittineni

(E82-10119) PROBABILISTIC CLUSTER LABELING OF IMAGERY DATA (Lockheed Engineering and Management) 44 p HC A03/MF A01 CSCL 05B

N82-22590

Unclas
G3/43 00119

Lockheed Engineering and Management Services Company, Inc.
1830 NASA Road 1, Houston, Texas 77058

Lyndon B. Johnson Space Center
Houston, Texas 77058

1. Report No. SR-LO-00483, JSC-16384	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle
Probabilistic Cluster Labeling of Imagery Data

5. Report Date
September 1980 6

6. Performing Organization Code

7. Author(s)
C. B. Chittineni
Lockheed Engineering and Management Services Comapny, Inc.

8. Performing Organization Report No.
LEMSCO-15358

9. Performing Organization Name and Address
Lockheed Engineering and Management Services Company, Inc.
1830 NASA Road 1
Houston, Texas 77058

10. Work Unit No.

11. Contract or Grant No.
NAS 9-15800 5

12. Sponsoring Agency Name and Address
National Aeronautics and Space Administration 1
Lyndon B. Johnson Space Center
Houston, Texas 77058 Technical Monitor: Dr. R. Heydorn

13. Type of Report and Period Covered
Technical Report

14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

In this paper the author considers the problem of obtaining the probabilities of class labels for the clusters using spectral and spatial information from a given set of labeled patterns and their neighbors. A relationship is developed between class and cluster conditional densities in terms of probabilities of class labels for the clusters. Expressions are presented for updating the a posteriori probabilities of the classes of a pixel using information from its local neighborhood. Fixed-point iteration schemes are developed for obtaining the optimal probabilities of class labels for the clusters. These schemes utilize spatial information and also the probabilities of label imperfections. Furthermore, experimental results from the processing of remotely sensed multispectral scanner imagery data are presented.

17. Key Words (Suggested by Author(s))
Analyst labels, cluster labeling, fixed-point iteration equation, maximum likelihood criterion, probabilities of class labels for the clusters, probabilities of label imperfections, remote sensing, spatial information, transition probabilities

18. Distribution Statement

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
Unclassified	Unclassified	42	

*For sale by the National Technical Information Service, Springfield, Virginia 22161

SR-L0-00483
JSC-16384

PROBABILISTIC CLUSTER LABELING OF IMAGERY DATA

Job Order 73-345

This report describes Classification activities of the Supporting Research project of the AgRISTARS program.

PREPARED BY

C. B. Chittineni

APPROVED BY

T. C. Minter

T. C. Minter, Supervisor
Techniques Development Section

J. E. Wainwright

J. E. Wainwright, Manager
Development and Evaluation Department

LOCKHEED ENGINEERING AND MANAGEMENT SERVICES COMPANY, INC.

Under Contract NAS 9-15800

For

Earth Observations Division

Space and Life Sciences Directorate

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
LYNDON B. JOHNSON SPACE CENTER
HOUSTON, TEXAS

September 1980

LEMSCO-15358

PREFACE

The techniques which are the subject of this report were developed to support the Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing program. Under Contract NAS 9-15800, Dr. C. B. Chittineni, a principal scientist for Lockheed Engineering and Management Services Company, Inc., performed this research for the Earth Observations Division, Space and Life Science Directorate, National Aeronautics and Space Administration, at the Lyndon B. Johnson Space Center, Houston, Texas

PRECEDING PAGE BLANK NOT FILMED

CONTENTS

TABLES

FIGURE

1. INTRODUCTION

Recently, considerable interest has been shown in developing techniques for the classification of imagery data such as remote sensing data obtained using the multispectral scanner (MSS) on board the Landsat for inventorying natural resources, monitoring crop conditions, detecting mineral and oil deposits, etc. Usually, the inherent classes in the data are multimodal, and nonsupervised classification or clustering techniques (refs. 1-3) have been found to be effective (refs. 4, 5) in the classification of imagery data. Clustering the data partitions the image into its inherent modes or clusters. Labeling the clusters is one of the crucial problems in the application of clustering techniques for the classification of imagery data.

Cluster labeling is similar to the problem of labeling the regions obtained by using segmentation algorithms in the development of scene understanding systems. The recent literature shows considerable interest in the use of relaxation labeling algorithms for labeling the segmented regions (refs. 6-8). These algorithms use relational properties of the regions through compatibility coefficients. In cluster labeling, the relational properties of the clusters are either not available or not meaningful. For example, in aerospace agricultural imagery, the regions of interest are crops, nonagricultural areas, etc. These can be anywhere in the image. Hence, it is not meaningful to define relational properties for the clusters.

Most of the imagery data contain much spatial information, and several researchers (refs. 9-12) have attempted to use spatial information in the classification of imagery data.

This paper documents an investigation of the problem of labeling the clusters using spectral and spatial information. It is assumed that the probability density functions and a priori probabilities of the clusters or modes are given. Let these respectively be $p(X|\Omega = i)$ and δ_i; $i = 1,2,\cdots,m$, where m is the number of modes or clusters. It is also assumed that a set of labeled patterns $X_i(j)$ with labels $\omega_i(j) = i$ and their neighboring patterns

$Y_i^k(j)(k = 1,2,\cdots,\ell;j = 1,2,\cdots,N_i$ and $i = 1,2,\cdots,C)$ are given, where C is the number of classes.

In remote sensing, the labels for the patterns are provided by an analyst interpreter (AI), who examines imagery films and uses other data such as historic information and crop calendar models. Very often the AI labels are imperfect. Recently, Chittineni (refs. 13-15) investigated techniques for the estimation of probabilities of label imperfections using imperfectly labeled and unlabeled patterns. It is assumed that the probabilities of label imperfections are available. Methods are developed in the paper for obtaining probabilities of class labels for the clusters using all the available information.

This paper is organized as follows. In section 2, a relationship is developed between class conditional densities and cluster conditional densities in terms of probabilities of class labels for the clusters. Section 3 concerns the problem of obtaining probabilities of class labels for the clusters without using spatial information. Expressions are presented in section 4 for updating the a posteriori probabilities of the classes of a pixel using spectral and spatial information from its neighborhood. Section 5 deals with the problem of obtaining probabilities of class labels for the clusters using spectral and spatial information. Imperfections in the labels of the given pattern set are considered in section 6. Section 7 contains the experimental results in the processing of remotely sensed imagery data, and the concluding remarks are given in section 8. In Appendix A, the problem of obtaining the probabilities of class labels for the clusters using information from a given set of labeled fields is considered. Contextual cluster labeling with the probability of correct labeling as a criterion is treated in Appendix B.

2

2. A RELATIONSHIP BETWEEN CLUSTER AND CLASS CONDITIONAL DENSITIES

In this section, a relationship is developed between cluster and class conditional densities. In general, the class conditional density functions are multimodal. Let C be the number of classes and m be the number of clusters. Let $p(X|\omega = i)$ be the class conditional densities and $p(X|\Omega = i)$ be the mode or cluster conditional densities. Let $P(\omega = i)$ and $P(\Omega = i)$ be the a priori probability of class i and the a priori probability of cluster i, respectively. The mixture density p(X) can be written in terms of class conditional densities as

$$p(X) = \sum_{i=1}^{C} P(\omega = i)p(X|\omega = i) \tag{1}$$

The mixture density p(X) can also be written in terms of mode conditional densities as

$$\begin{aligned} p(X) &= \sum_{\ell=1}^{m} P(\Omega = \ell)p(X|\Omega = \ell) \\ &= \sum_{\ell=1}^{m} p(X|\Omega = \ell) \sum_{i=1}^{C} P(\Omega = \ell, \omega = i) \\ &= \sum_{i=1}^{C} P(\omega = i) \sum_{\ell=1}^{m} P(\Omega = \ell|\omega = i)p(X|\Omega = \ell) \end{aligned} \tag{2}$$

The following assumption is made from comparing equations (1) and (2).

$$p(X|\omega = i) = \sum_{\ell=1}^{m} P(\Omega = \ell|\omega = i)p(X|\Omega = \ell) \tag{3}$$

Equation (3) can be rewritten as

$$p(\omega = i|X) = \sum_{\ell=1}^{m} \alpha_{\ell i}p(\Omega = \ell|X) \tag{4}$$

where $\alpha_{\ell i} = P(\omega = i|\Omega = \ell)$ and is the probability that the label of mode ℓ is class i. The probabilities $\alpha_{\ell i}$ satisfy the constraints given in equation (5).

$$\alpha_{\ell i} \geq 0 \quad ; \quad i = 1,2,\cdots,C \text{ and } \ell = 1,2,\cdots,m$$

$$\sum_{i=1}^{C} \alpha_{\ell i} = 1 \quad ; \quad \ell = 1,2,\cdots,m \tag{5}$$

Equation (3) provides a relationship between class and cluster conditional densities in terms of probabilities of class labels for the clusters.

4

3. MAXIMUM LIKELIHOOD PROBABILISTIC CLUSTER LABELING

This section concerns the problem of obtaining the probabilities $\alpha_{\ell i}$ (the probabilities of class labels for the clusters). It is assumed that we are given a set of labeled patterns $X_i(j)$ with class labels $\omega_i(j) = i$; $j = 1,2,\cdots,N_i$ and $i = 1,2,\cdots,C$. It is also assumed that the a priori probabilities of the modes or clusters and mode conditional densities are given. Let δ_i and $p(X|\Omega = i)$ be the mode a priori probabilities and mode conditional densities, respectively. The criterion used in obtaining the probabilistic description of class labels for the clusters is the likelihood function. The likelihood of an occurrence of patterns $X_i(j)$ with their labels $\omega_i(j) = i$ is given by

$$L_1' = \prod_{i=1}^{C} \prod_{j=1}^{N_i} p[X_i(j), \omega_i(j) = i] \tag{6}$$

Since $\prod_{i=1}^{C} \prod_{j=1}^{N_i} p[X_i(j)]$ is independent of $\omega_i(j)$, for mathematical simplicity, dividing the above equation by it yields

$$L_1 = \prod_{i=1}^{C} \prod_{j=1}^{N_i} \frac{p[X_i(j), \omega_i(j) = i]}{p[X_i(j)]} \tag{7}$$

Noting that the logarithm is a monotonic function of its argument and taking the logarithm of L_1 of equation (7) and using equation (4) yield the following.

$$L = \log(L_1) = \sum_{i=1}^{C} \sum_{j=1}^{N_i} \log\left\{\sum_{\ell=1}^{m} \alpha_{\ell i} p[\Omega = \ell | X_i(j)]\right\} \tag{8}$$

The probabilities $\alpha_{\ell i}$ satisfy the constraints given in equation (5). Closed-form solutions for $\alpha_{\ell i}$ by maximizing L of equation (8), subject to the constraints of equation (5), seem to be difficult to obtain. The probabilities $\alpha_{\ell i}$ can easily be obtained using optimization techniques such as the Davidon-Fletcher-Powell procedure (refs. 16-18).

3.1 A FIXED-POINT ITERATION SCHEME FOR OPTIMAL $\alpha_{\ell i}$

The following fixed-point iteration equation (similar to maximum likelihood equations in parametric clustering in reference 3) for the solution of the above optimization problem can easily be obtained by introducing Lagrangian multipliers. That is:

$$\alpha_{\ell i} = \frac{\sum_{j=1}^{N_i} d_{\ell ij}}{\sum_{i=1}^{C} \sum_{j=1}^{N_i} d_{\ell ij}} \tag{9}$$

where
$$d_{\ell ij} = \frac{\alpha_{\ell i} p[\Omega = \ell | X_i(j)]}{\sum_{s=1}^{m} \alpha_{si} p[\Omega = s | X_i(j)]} \tag{10}$$

However, closed-form solutions for $\alpha_{\ell i}$ can be obtained with the criterion as the maximization of a lower bound on L, and they are given in the next section.

3.2 CLOSED-FORM SOLUTIONS FOR THE PROBABILITIES $\alpha_{\ell i}$

Since the logarithm is a convex upward function, we have the inequality

$$\log\left[\sum_{i=1}^{C} a_i g_i(X)\right] > \sum_{i=1}^{C} a_i \log[g_i(X)] \tag{11}$$

where
$$\sum_{i=1}^{C} a_i = 1$$

and
$$a_i > 0 \quad ; \quad i = 1,2,\cdots,C \tag{12}$$

Using the inequality of equation (11) in equation (8), a lower bound on the log likelihood function L can be obtained as

$$L > \sum_{i=1}^{C} \sum_{j=1}^{N_i} \sum_{\ell=1}^{m} \left\{ p[\Omega = \ell | X_i(j)] \log(\alpha_{\ell i}) \right\} \tag{13}$$

6

With the introduction of the Lagrangian multipliers, the probabilities $\alpha_{\ell i}$ that maximize the lower bound of equation (13), subject to the constraints of equation (5), can be obtained as follows.

$$\alpha_{\ell i} = \frac{N_i e_{i\ell}}{\sum_{r=1}^{C} N_r e_{r\ell}} \tag{14}$$

where
$$e_{i\ell} = \frac{1}{N_i} \sum_{j=1}^{N_i} p[\Omega = \ell | X_i(j)] \tag{15}$$

This solution simply states that the probability of the i^{th} class label for a given cluster ℓ is the ratio of the sum of the a posteriori probabilities of cluster ℓ given the labeled patterns from class i to the sum over all classes of the sum of a posteriori probabilities of cluster ℓ given the labeled patterns from each class. Having obtained $\alpha_{\ell i}$, q_i (the proportion of class i) can be estimated as follows.

$$\begin{aligned} q_i &= P(\omega = i) \\ &= \sum_{\ell=1}^{m} P(\omega = i, \Omega = \ell) = \sum_{\ell=1}^{m} \delta_\ell \alpha_{\ell i} \end{aligned} \tag{16}$$

Hence, $\hat{q}_i$ (the estimate of q_i) can be computed from the following.

$$\hat{q}_i = \sum_{\ell=1}^{m} \delta_\ell \hat{\alpha}_{\ell i} \tag{17}$$

4. UPDATING A POSTERIORI PROBABILITIES OF THE CLASSES OF A PIXEL USING INFORMATION FROM ITS NEIGHBORHOOD

The last section covered the problem of estimating the probabilities $\alpha_{\ell i}$ (the probabilities of class labels for the clusters) using information from a given set of labeled patterns. The probabilities $\alpha_{\ell i}$ are seen to be functions of $p[\omega_i(j) = i|X_i(j)]$;the a posteriori probabilities of the classes and spatial information is not used in obtaining $\alpha_{\ell i}$. Most of the natural imagery is abundant in spatial information and can be used to obtain better estimates for $\alpha_{\ell i}$. In this section, expressions are developed for updating the a posteriori probabilities of the classes of a picture element (pixel) using information from its local neighborhood. These expressions are used in section 5 to obtain the probabilities of class labels for the clusters using both the spectral and spatial information.

Let the pixel under consideration be pixel 0. Its four neighbors in a two-dimensional local neighborhood are shown in figure 1.

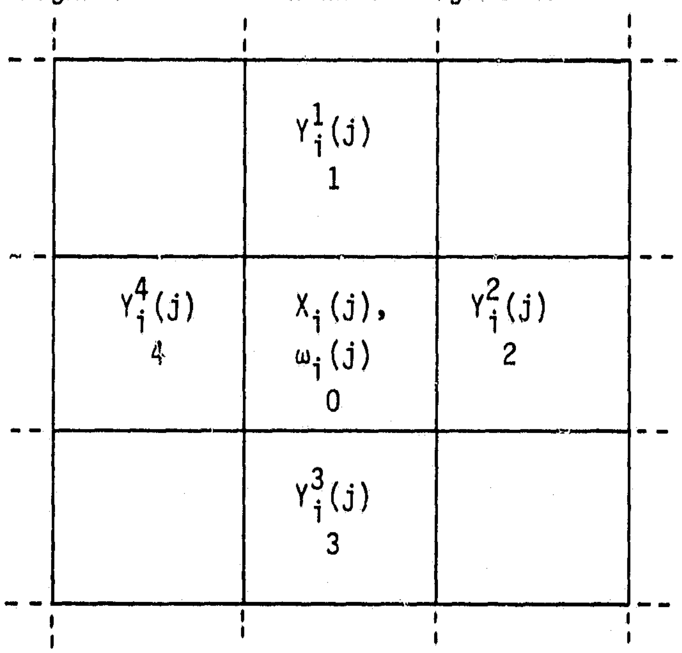

Figure 1.- Four neighbors of a pixel 0.

8

The following a posteriori probabilities of the classes of a pixel 0 are obtained by using information from its local neighborhood.

$$p\left[\omega_i(j) = k | X_i(j), Y_i^1(j), \cdots, Y_i^4(j)\right] = \frac{p\left[\omega_i(j) = k, X_i(j), Y_i^1(j), \cdots, Y_i^4(j)\right]}{p\left[X_i(j), Y_i^1(j), \cdots, Y_i^4(j)\right]} \tag{18}$$

The denominator of equation (18) can be written as

$$p\left[X_i(j), Y_i^1(j), \cdots, Y_i^4(j)\right] = \sum_{k=1}^{C} p\left[\omega_i(j) = k, X_i(j), Y_i^1(j), \cdots, Y_i^4(j)\right] \tag{19}$$

Similarly, from the numerator of equation (18), we obtain

$$\begin{aligned} & p\left[\omega_i(j) = k, X_i(j), Y_i^1(j), \cdots, Y_i^4(j)\right] \\ & = \sum_{k_1=1}^{C} \cdots \sum_{k_4=1}^{C} p\left[\omega_i(j) = k, X_i(j), \omega_i^1(j) = k_1, Y_i^1(j), \cdots, \omega_i^4(j) = k_4, Y_i^4(j)\right] \\ & = \sum_{k_1=1}^{C} \cdots \sum_{k_4=1}^{C} p\left[X_i(j), Y_i^1(j), \cdots, Y_i^4(j) | \omega_i(j) = k, \omega_i^1(j) = k_1, \cdots, \omega_i^4(j) = k_4\right] \\ & \quad p\left[\omega_i(j) = k, \omega_i^1(j) = k_1, \cdots, \omega_i^4(j) = k_4\right] \end{aligned} \tag{20}$$

where C is the number of classes.

In the following, it is assumed (a) that the probability density function of a pattern, given its label, is independent of other patterns and their labels and (b) that the labels of the patterns are independent of the labels of their nonneighbors. In the following analysis, the pixels having a common side are considered as neighbors. (For example, in figure 1, pixels 0 and 1 are neighbors, whereas pixels 1 and 2 are nonneighbors.) By repeatedly using assumption (a), the following is obtained.

9

$$p\left[X_i(j),Y_i^1(j),\cdots,Y_i^4(j)|\omega_i(j) = k,\omega_i^1(j) = k_1,\cdots,\omega_i^4(j) = k_4\right]$$

$$= p\left[X_i(j)|\omega_i(j) = k,Y_i^1(j),\omega_i^1(j) = k_1,\cdots,Y_i^4(j),\omega_i^4(j) = k_4\right]$$

$$p\left[Y_i^1(j),\cdots,Y_i^4(j)|\omega_i(j) = k,\omega_i^1(j) = k_1,\cdots,\omega_i^4(j)\right] = k_4$$

$$= p[X_i(j)|\omega_i(j) = k]p\left[Y_i^1(j),\cdots,Y_i^4(j)|\omega_i(j) = k,\omega_i^1(j) = k_1,\cdots,\omega_i^4(j) = k_4\right]$$

$$= p[X_i(j)|\omega_i(j) = k]\left\{\prod_{\ell=1}^{4} p\left[Y_i^\ell(j)|\omega_i^\ell(j) = k_\ell\right]\right\} \tag{21}$$

By repeatedly using assumption (b), the second term in the summations of equation (20) can be written as follows.

$$P\left[\omega_i(j) = k,\omega_i^1(j) = k_1,\cdots,\omega_i^4(j) = k_4\right]$$

$$= P[\omega_i(j) = k]P\left[\omega_i^1(j) = k_1,\cdots,\omega_i^4(j) = k_4|\omega_i(j) = k\right]$$

$$= P[\omega_i(j) = k]P\left[\omega_i^1(j) = k_1|\omega_i(j) = k,\omega_i^2(j) = k_2,\cdots,\omega_i^4(j) = k_4\right]$$

$$P\left[\omega_i^2(j) = k_2,\cdots,\omega_i^4(j) = k_4|\omega_i(j) = k\right]$$

$$= P[\omega_i(j) = k]\left\{\prod_{\ell=1}^{4} P\left[\omega_i^\ell(j) = k_\ell|\omega_i(j) = k\right]\right\} \tag{22}$$

Using equations (21) and (22) in equation (20) results in

$$P\left[\omega_i(j) = k, X_i(j), Y_i^1(j), \cdots, Y_i^4(j)\right]$$

$$= P[\omega_i(j) = k]\, p[X_i(j) | \omega_i(j) = k] \sum_{k_1=1}^{C} \cdots \sum_{k_4=1}^{C}$$

$$\left(\prod_{\ell=1}^{4} \left\{ P\left[\omega_i^\ell(j) = k_\ell | \omega_i(j) = k\right] p\left[Y_i^\ell(j) | \omega_i^\ell(j) = k_\ell\right] \right\} \right)$$

$$= P[\omega_i(j) = k]\, p[X_i(j) | \omega_i(j) = k]$$

$$\prod_{\ell=1}^{4} \left(\sum_{k_\ell=1}^{C} \left\{ P\left[\omega_i^\ell(j) = k_\ell | \omega_i(j) = k\right] p\left[Y_i^\ell(j) | \omega_i^\ell(j) = k_\ell\right] \right\} \right) \tag{23}$$

From equations (18), (19), and (23), we obtain

$$p\left[\omega_i(j) = k | X_i(j), Y_i^1(j), \cdots, Y_i^4(j)\right]$$

$$= \frac{p[\omega_i(j) = k | X_i(j)] \displaystyle\prod_{\ell=1}^{4} \left(\sum_{k_\ell=1}^{C} \left\{ \frac{P\left[\omega_i^\ell(j) = k_\ell | \omega_i(j) = k\right]}{P\left[\omega_i^\ell(j) = k_\ell\right]} p\left[\omega_i^\ell(j) = k_\ell | Y_i^\ell(j)\right] \right\} \right)}{\displaystyle\sum_{k=1}^{C} p[\omega_i(j) = k | X_i(j)] \prod_{\ell=1}^{4} \left(\sum_{k_\ell=1}^{C} \left\{ \frac{P\left[\omega_i^\ell(j) = k_\ell | \omega_i(j) = k\right]}{P\left[\omega_i^\ell(j) = k_\ell\right]} p\left[\omega_i^\ell(j) = k_\ell | Y_i^\ell(j)\right] \right\} \right)} \tag{24}$$

In equation (24), the spectral and spatial information from the neighborhood of a pixel is used in obtaining the a posteriori probabilities of its classes.

5. PROBABILISTIC CLUSTER LABELING WITH SPECTRAL AND SPATIAL INFORMATION

This section covers the problem of obtaining the probabilities $\alpha_{\ell i}$ (the probabilities of class labels for the clusters) using spectral and spatial information. It is assumed that we are given a set of labeled patterns $X_i(j)$ with labels $\omega_i(j) = i$ and their neighbors $Y_i^{\ell}(j)$, $\ell = 1,2,\cdots,4$, as shown in figure 1. For $j = 1,2,\cdots, N_i$ and, $i = 1,2,\cdots,C$, the likelihood of occurrence of patterns $X_i(j)$ with labels $\omega_i(j) = i$ and with $Y_i^{\ell}(j)$, $\ell = 1,2,\cdots,4$, as their neighbors is given as

$$L' = \prod_{i=1}^{C} \prod_{j=1}^{N_i} \left\{ p\left[X_i(j), \omega_i(j) = i, Y_i^1(j), \cdots, Y_i^4(j)\right]\right\} \tag{25}$$

From equations (24) and (25), the log likelihood function can be written as

$$L = \sum_{i=1}^{C} \sum_{j=1}^{N_i} \log\{p[\omega_i(j) = i | X_i(j)]\}$$

$$+ \sum_{i=1}^{C} \sum_{j=1}^{N_i} \sum_{\ell=1}^{4} \log\left(\sum_{i_\ell=1}^{C} \left\{ p\left[\omega_i^{\ell}(j) = i_\ell | Y_i^{\ell}(j)\right] \frac{p\left[\omega_i^{\ell}(j) = i_\ell | \omega_i(j) = i\right]}{p\left[\omega_i^{\ell}(j) = i_\ell\right]}\right\}\right) \tag{26}$$

Using equation (4) in equation (26) yields

$$L = \sum_{i=1}^{C} \sum_{j=1}^{N_i} \log\left\{\sum_{r=1}^{m} \alpha_{ri} p[\Omega = r | X_i(j)]\right\}$$

$$+ \sum_{i=1}^{C} \sum_{j=1}^{N_i} \sum_{\ell=1}^{4} \log\left(\sum_{i_\ell=1}^{C} \frac{p\left[\omega_i^{\ell}(j) = i_\ell | \omega_i(j) = i\right]}{p\left[\omega_i^{\ell}(j) = i_\ell\right]} \left\{\sum_{r=1}^{m} \alpha_{ri_\ell} p\left[\Omega = r | Y_i^{\ell}(j)\right]\right\}\right) \tag{27}$$

12

Closed-form solutions for the probabilities α_{ri} that maximize L of equation (27), subject to the constraints of equation (5), seem to be difficult. Optimization methods (refs. 16-18) such as the Davidon-Fletcher-Powell procedure can easily be used to obtain probabilities α_{ri} that maximize L of equation (27), subject to the constraints of equation (5). By introducing Lagrangian multipliers, the following fixed-point iteration equation for the solution of the above optimization problem can easily be obtained. That is,

$$\alpha_{ri} = \frac{\alpha_{ri}\left(\left\{\sum_{j=1}^{N_i} \frac{p[\Omega = r|X_i(j)]}{\sum_{s=1}^{m} \alpha_{si} p[\Omega = s|X_i(j)]}\right\} + \delta_{ri}\right)}{\sum_{i=1}^{C} \alpha_{ri}\left(\left\{\sum_{j=1}^{N_i} \frac{p[\Omega = r|X_i(j)]}{\sum_{s=1}^{m} \alpha_{si} p[\Omega = s|X_i(j)]}\right\} + \delta_{ri}\right)} \tag{28}$$

where

$$\delta_{ri} = \sum_{k=1}^{C} \sum_{j=1}^{N_k} \sum_{\ell=1}^{4} \frac{\dfrac{P\left[\omega_k^\ell(j) = i|\omega_k(j) = k\right]}{P\left[\omega_k^\ell(j) = i\right]} p\left[\Omega = r|Y_k^\ell(j)\right]}{\sum_{i_\ell=1}^{C}\left(\left\{\dfrac{P\left[\omega_k^\ell(j) = i_\ell|\omega_k(j) = k\right]}{P\left[\omega_k^\ell(j) = i_\ell\right]}\right\} \sum_{s=1}^{m} \alpha_{si_\ell} p\left[\Omega = s|Y_k^\ell(j)\right]\right)} \tag{29}$$

If the spatial information is not used (that is, when $\delta_{ri} = 0$), it is easily seen that equation (28) becomes identical to equation (9).

6. CLUSTER LABELING WHEN THE LABELS OF THE GIVEN PATTERN SET ARE IMPERFECT

In practice, such as in the classification of remotely sensed, MSS imagery data, it is difficult and expensive to obtain labels for the training patterns. The labels for the patterns are usually provided by an AI who examines imagery films and uses some other information. (For example, in labeling pixels of remote sensing agricultural imagery, the information that is most often used is historic information, crop growth stage models, etc.) These labels are very often imperfect. Recently, there has been considerable interest (refs. 13-15) in estimating the probabilities of label imperfections and using these estimates to obtain the improved classification and to identify mislabeled patterns with a specified degree of confidence. This section pertains to the problem of probabilistic cluster labeling by taking into acount the imperfections in the labels of the given labeled pattern set. Let ω and ω' be the perfect and imperfect labels, respectively, each of which takes values $1,2,\cdots,C$. The imperfections in the labels are described by the probabilities

$$\beta_{ji} = P(\omega' = i|\omega = j) \tag{30}$$

where

$$\sum_{i=1}^{C} \beta_{ji} = 1 \tag{31}$$

To obtain a relationship between class conditional densities with and without imperfections in the labels, consider

$$\begin{aligned} p(X|\omega' = i) &= \frac{1}{P(\omega' = i)} \sum_{j=1}^{C} p(X,\omega' = i,\omega = j) \\ &= \frac{1}{P(\omega' = i)} \sum_{j=1}^{C} p(X|\omega' = i,\omega = j)P(\omega' = i|\omega = j)P(\omega = j) \\ &= \frac{1}{P(\omega' = i)} \sum_{j=1}^{C} \beta_{ji}P(\omega = j)p(X|\omega = j) \end{aligned} \tag{32}$$

14

where it is assumed that

$$p(X|\omega' = i, \omega = j) = p(X|\omega = j) \tag{33}$$

Using the Bayes rule, from equation (32), we obtain

$$p(\omega' = i|X) = \sum_{j=1}^{C} \beta_{ji} p(\omega = j|X) \tag{34}$$

In the following, it is assumed that a set of labeled patterns $X_i(j)$ with imperfect labels $\omega_i'(j) = i$ and with the neighbors $Y_i^1(j), \cdots, Y_i^4(j)$ as shown in figure 1 for $j = 1, 2, \cdots, N_i$ and $i = 1, 2, \cdots, C$ is given. It is also assumed that the probabilities of label imperfections β_{ji} are available. The probabilities of imperfections in the labels being β_{ji}, the likelihood of the occurrence of patterns $X_i(j)$ with imperfect labels $\omega_i'(j) = i$, $Y_i^1(j), \cdots, Y_i^4(j)$ being their neighbors is given by the following

$$L' = \prod_{i=1}^{C} \prod_{j=1}^{N_i} p\left[\omega_i'(j) = i, X_i(j), Y_i^1(j), \cdots, Y_i^4(j)\right] \tag{35}$$

Consider

$$\begin{aligned} p\left[\omega_i'(j) = k, X_i(j), Y_i^1(j), \cdots, Y_i^4(j)\right] &= \sum_{k_1=1}^{C} \cdots \sum_{k_4=1}^{C} p\left[\omega_i'(j) = k, X_i(j), Y_i^1(j), \omega_i^1(j) = k_1, \cdots, Y_i^4(j), \omega_i^4(j) = k_4\right] \\ &= \sum_{k_1=1}^{C} \cdots \sum_{k_4=1}^{C} p\left[X_i(j), Y_i^1(j), \cdots, Y_i^4(j) | \omega_i'(j) = k, \omega_i^1(j) = k_1, \cdots, \omega_i^4(j) = k_4\right] \\ &\quad p\left[\omega_i'(j) = k, \omega_i^1(j) = k_1, \cdots, \omega_i^4(j) = k_4\right] \end{aligned} \tag{36}$$

Given the probabilities of imperfections in the labels and proceeding similarly to equations (21) and (22) while using assumptions (a) and (b) of section 4, the following can easily be obtained.

$$p\left[X_i(j), Y_i^1(j), \cdots, Y_i^4(j) \mid \omega_i'(j) = k, \omega_i^1(j) = k_1, \cdots, \omega_i^4(j) = k_4\right]$$

$$= p[X_i(j) \mid \omega_i'(j) = k] \left\{ \prod_{\ell=1}^{4} p\left[Y_i^\ell(j) \mid \omega_i^\ell(j) = k_\ell\right] \right\} \tag{37}$$

and

$$P\left[\omega_i'(j) = k, \omega_i^1(j) = k_1, \cdots, \omega_i^4(j) = k_4\right] = P[\omega_i'(j) = k]$$

$$\left\{ \prod_{\ell=1}^{4} P\left[\omega_i^\ell(j) = k_\ell \mid \omega_i'(j) = k\right] \right\} \tag{38}$$

Since $p[X_i(j)]\left\{\prod_{\ell=1}^{4} p\left[Y_i^\ell(j)\right]\right\}$ is a constant, dividing equation (35) by it and using equations (37) and (38) in equation (35) yields

$$\frac{p\left[\omega_i'(j) = k, X_i(j), Y_i^1(j), \cdots, Y_i^4(j)\right]}{p[X_i(j)]\left\{\prod_{\ell=1}^{4} p\ Y_i^\ell(j)\right\}}$$

$$= p[\omega_i'(j) = k \mid X_i(j)] \prod_{\ell=1}^{4} \left(\sum_{k_\ell=1}^{C} p\left[\omega_i^\ell(j) = k_\ell \mid Y_i^\ell(j)\right] \left\{ \frac{P\left[\omega_i^\ell(j) = k_\ell \mid \omega_i'(j) = k\right]}{P\ \omega_i^\ell(j) = k_\ell} \right\} \right)$$

$$= \frac{p[\omega_i'(j) = k \mid X_i(j)]}{P\left[\omega_i'(j) = k\right]} \prod_{\ell=1}^{4} \left\{ \sum_{k_\ell=1}^{C} p\left[\omega_i^\ell(j) = k_\ell \mid Y_i^\ell(j)\right] \nu_{k_\ell i} \right\} \tag{39}$$

where $$\nu_{k_\ell i} = \sum_{s=1}^{C} \beta_{si} \frac{P[\omega_i(j) = s]}{P\left[\omega_i^\ell(j) = k_\ell\right]} P\left[\omega_i^\ell(j) = k_\ell \mid \omega_i(j) = s\right] \tag{40}$$

and it is assumed that

$$\beta_{sk} = P\left[\omega_i'(j) = k \middle| \omega_i(j) = s\right]$$
$$= P\left[\omega_i'(j) = k \middle| \omega_i(j) = s, \omega_i^{\ell}(j) = k_{\ell}\right] \tag{41}$$

Since the logarithm is a monotonic function of its argument, taking the logarithm of equation (35), using equation (39) in equation (35), and treating a priori probabilities of the imperfect labels as constant, the log likelihood function becomes

$$L = \sum_{i=1}^{C} \sum_{j=1}^{N_i} \log\left\{p\left[\omega_i'(j) = i \middle| X_i(j)\right]\right\}$$

$$+ \sum_{i=1}^{C} \sum_{j=1}^{N_i} \sum_{\ell=1}^{4} \log\left\{\sum_{k_{\ell}=1}^{C} \nu_{k_{\ell} i}\, p\left[\omega_i^{\ell}(j) = k_{\ell} \middle| Y_i^{\ell}(j)\right]\right\} \tag{42}$$

Using equations (4) and (33) in equation (42) yields

$$L = \sum_{i=1}^{C} \sum_{j=1}^{N_i} \log\left\{\sum_{r=1}^{m} \sum_{s=1}^{C} \alpha_{rs} \beta_{si}\, p[\Omega = r | X_i(j)]\right\}$$

$$+ \sum_{i=1}^{C} \sum_{j=1}^{N_i} \sum_{\ell=1}^{4} \log\left\{\sum_{r=1}^{m} \sum_{k_{\ell}=1}^{C} \alpha_{rk_{\ell}} \nu_{k_{\ell} i}\, p\left[\Omega = r \middle| Y_i^{\ell}(j)\right]\right\} \tag{43}$$

Optimization methods such as the Davidon-Fletcher-Powell procedure (refs. 16-18) can easily be used to the obtain optimal α_{uv} that maximizes L, subject to the constaints of equation (5). Also, fixed-point iteration equations similar to equation (28) can easily be derived to obtain the optimal α_{uv} by introducing Lagrangian multipliers and are given in the following.

$$\alpha_{uv} = \frac{\alpha_{uv}\left(\delta_{uv}^{1} + \delta_{uv}^{2}\right)}{\sum_{s=1}^{C} \alpha_{us}\left(\delta_{us}^{1} + \delta_{us}^{2}\right)} \tag{44}$$

17

where

$$\delta_{uv}^{1} = \sum_{i=1}^{C} \sum_{j=1}^{N_i} \left\{ \frac{\beta_{vi} p[\Omega = u|X_i(j)]}{\sum_{r=1}^{m} \sum_{s=1}^{C} \alpha_{rs} \beta_{si} p[\Omega = r|X_i(j)]} \right\} \tag{45}$$

and

$$\delta_{uv}^{2} = \sum_{i=1}^{C} \sum_{j=1}^{N_i} \sum_{\ell=1}^{4} \frac{\nu_{vi} p\left[\Omega = u|Y_i^{\ell}(j)\right]}{\sum_{r=1}^{m} \sum_{k_\ell=1}^{C} \alpha_{rk_\ell} \nu_{k_\ell i} p\left[\Omega = r|Y_i^{\ell}(j)\right]} \tag{46}$$

If the spatial information is not used, the fixed-point iteration equation (44) for obtaining α_{uv}, the probabilities of class labels for the clusters become the following.

$$\alpha_{uv} = \frac{\sum_{i=1}^{C} \sum_{j=1}^{N_i} d_{ijvu}}{\sum_{v=1}^{C} \sum_{i=1}^{C} \sum_{j=1}^{N_i} d_{ijvu}} \tag{47}$$

where

$$d_{ijvu} = \frac{\alpha_{uv} \beta_{vi} p[\Omega = u|X_i(j)]}{\sum_{s=1}^{C} \sum_{k=1}^{m} \alpha_{ks} \beta_{si} p[\Omega = k|X_i(j)]} \tag{48}$$

7. EXPERIMENTAL RESULTS

This section presents some results obtained in the processing of remotely sensed Landsat MSS imagery data. The objective of the processing is to estimate the proportion of the class of interest in each image. There are two classes in the image. Class 1 is wheat, and class 2 is nonwheat, which is designated as "other." The class of interest is wheat. The MSS images of several segments were processed in the following manner. [A segment is a 9- by 11-kilometer (5- by 6-nautical-mile) area for which the MSS image is divided into a rectangular array of pixels, 117 rows by 196 columns.] The image is overlaid with a rectangular grid of 209 grid intersections.

Class labels were given to the pixels corresponding to a subset of 209 grid intersections by an AI who examined the imagery films and used some other information such as crop growth stage models and historic information. These are imperfect labels. Also, ground-truth labels or true labels of these pixels are acquired.

The numbers and locations of the segments, the number of pixels labeled, and the number of features or the number of channels used for each segment are listed in table 1. Several acquisitions were used for each segment. The Gaussian mode (cluster) conditional densities and a priori probabilities of the inherent modes in the data of each segment are obtained using a maximum likelihood clustering algorithm (refs. 3, 19). The number of clusters generated for each segment is listed in table 1. The theory developed in sections 3 and 5 is applied in estimating the probabilities of class labels for the clusters of each segment using AI-labeled patterns and ground-truth-labeled patterns, both with and without the use of contextual information.

The proportion of class 1, the class of interest, is estimated for each segment using equation (17) for all the cases, and the estimates are listed in table 1. The proportion of class 1 of each segment based on true [ground truth (GT)] labels of all the pixels in the segment is listed in the last column of table 1. In equations (28) and (29), the following a priori and

TABLE 1.- PROPORTION ESTIMATION THROUGH CLUSTER LABELING WITH AND WITHOUT THE USE OF CONTEXTUAL INFORMATION

Segment	Location (county, state)	No. of labeled patterns	Without context				With context		No. of features (d) and No. of clusters (m)	GT proportion
			AI labels		GT labels		AI labels	GT labels		
			Closed-form solution eq. (14)	Iterative solution eq. (9)	Closed-form solution eq. (14)	Iterative solution eq. (9)	Eq. (28)	Eq. (28)		
1005	Sherman, Texas	97	0.2311	0.2456	0.3621	0.3885	0.2430	0.3788	d = 8 m = 11	0.348
1060	Cheyenne, Colorado	106	0.1968	0.1975	0.3169	0.3251	0.1983	0.2863	d = 4 m = 11	0.231
1231	Jackson, Oklahoma	96	0.6378	0.6265	0.7395	0.7156	0.7066	0.7398	d = 6 m = 12	0.744
1520	Big Stone, Montana	91	0.2300	0.2109	0.2733	0.2696	0.2048	0.2543	d = 6 m = 10	0.301
1604	Renville, North Dakota	101	0.3214	0.2963	0.5023	0.4982	0.2690	0.4981	d = 4 m = 7	0.524
1675	McPherson, South Dakota	107	0.1019	0.1085	0.2977	0.3037	0.07641	0.2657	d = 8 m = 12	0.291
1805	Gregory, South Dakota	144	0.1156	0.1181	0.1649	0.1896	0.09073	0.1436	d = 8 m = 12	0.164
1853	Ness, Kansas	91	0.3161	0.3246	0.3200	0.3379	0.3043	0.3124	d = 6 m = 12	0.306
Bias			0.9479E-01	0.97625E-01	-0.8463E-02	-0.149E-01	0.10199	0.375E-02		
Mean square error			0.1380E-01	0.15142E-01	0.1135E-02	0.1849E-02	0.174E-01	0.9966E-03		

transition probabilities are used, where ω_C is the label of the central pixel and ω_N is the label of its neighbor.

$$\left.\begin{aligned} P(\omega_C = i) &= 0.5 \quad ; \quad i = 1,2 \\ P(\omega_N = j/\omega_C = i) &= \begin{cases} 0.8 \text{ if } i = j \\ 0.2 \text{ if } i \neq j \end{cases} \end{aligned}\right\} \tag{49}$$

From table 1, it is seen that considerable improvement has been made in the proportion estimates with the use of contextual information if the labels are good.

The probabilities of label imperfections of AI labels or the β-matrix are estimated for each segment by comparing imperfect (AI) labels and perfect (ground-truth) labels. These are listed in table 2. From tables 1 and 2, it is observed that, when the imperfections in the labels are small, the use of contextual information with the AI labels resulted in improved proportion estimates (see segment 1231).

Equations (44), (45), and (46) are used with the AI labels and the corresponding β-matrix for estimating the probabilities of class labels for the clusters. The values used for a priori and transition probabilities in these equations are given in equation (49). The resulting proportion estimates are listed in column 5 of table 2. The proportion of wheat in each segment is also estimated using equations (47) and (48) with the AI labels and the corresponding β-matrix. The resulting proportion estimates are listed in column 6 of table 2. From table 2, it is seen that there is considerable improvement in proportion estimates when the probabilities of label imperfections are taken into account.

TABLE 2.- ESTIMATION OF PROPORTION OF CLASS 1 WITH AI LABELS AND β-MATRIX

Segment	Location (county, state)	No. of AI-labeled patterns		Computed β-matrix comparing AI and GT labels	Proportion estimate using eqs. (44), (45), and (46)	Proportion estimate using eqs. (47) and (48)	Proportion estimate directly with eq. (9)	GT proportion
		Wheat	"Other"					
1005	Sherman, Texas	20	77	$\begin{bmatrix} 0.5455 & 0.4545 \\ 0.0308 & 0.9692 \end{bmatrix}$	0.3227	0.3025	0.2456	0.348
1060	Cheyenne, Colorado	17	89	$\begin{bmatrix} 0.5667 & 0.4343 \\ 0.0263 & 0.9737 \end{bmatrix}$	0.2174	0.2172	0.1975	0.231
1231	Jackson, Oklahoma	71	25	$\begin{bmatrix} 0.9315 & 0.0685 \\ 0.1304 & 0.8696 \end{bmatrix}$	0.6921	0.7139	0.6265	0.744
1520	Big Stone, Montana	20	71	$\begin{bmatrix} 0.7917 & 0.2083 \\ 0.0149 & 0.9851 \end{bmatrix}$	0.2432	0.3647	0.2109	0.301
1604	Renville, North Dakota	31	70	$\begin{bmatrix} 0.4600 & 0.5400 \\ 0.1569 & 0.8431 \end{bmatrix}$	0.4937	0.4814	0.2963	0.524
1675	Mcpherson, South Dakota	10	97	$\begin{bmatrix} 0.2667 & 0.7333 \\ 0.0390 & 0.9610 \end{bmatrix}$	0.2142	0.2156	0.1085	0.291
1805	Gregory, South Dakota	15	129	$\begin{bmatrix} 0.4211 & 0.5789 \\ 0.0640 & 0.9360 \end{bmatrix}$	0.1569	0.1932	0.1181	0.164
1853	Ness, Kansas	24	67	$\begin{bmatrix} 0.8077 & 0.1923 \\ 0.0615 & 0.9385 \end{bmatrix}$	0.3021	0.3052	0.3246	0.306
Bias					0.3271E-01	0.1441E-01	0.9763E-01	
Mean square error					0.1682E-02	0.1947E-02	0.1514E-01	

8. CONCLUDING REMARKS

In the classification of imagery data such as in the machine processing of remotely sensed MSS data, unsupervised classification techniques have been found to be effective. Clustering of the data partitions the image into its inherent modes. Labeling these clusters is one of the crucial problems in the application of clustering techniques for the classification of imagery data.

In the analysis of remotely sensed data, labels for the training patterns are usually provided by an AI who examines the imagery films and uses ancillary information such as historic information and crop growth stage models. These labels are usually imperfect. Most of the imagery data are abundant in spatial content, and spatial information improves the classification by machine processing.

In this paper, the problem of obtaining the probabilities of class labels for the clusters is considered. It is assumed that a set of labeled patterns $X_i(j)$ with class labels $\omega_i(j) = i$ and their neighbors $Y_i^{\ell}(j)(\ell = 1,2, \quad ,4;$ $j = 1,2, \quad ,N_i;$ and $i = 1,2, \quad ,C)$ are given, where C is the number of classes. The probabilities of imperfections in the labels are assumed to be available. It is also assumed that the number of inherent modes in the data, mode conditional densities, and a priori probabilities of the modes are given. Expressions are developed for obtaining the probabilities of class labels for the clusters using all the available information.

Experimental results are obtained from the processing of remotely sensed MSS imagery data. One of the important objectives in the analysis of remotely sensed data is to estimate the proportion of the crop of interest. In estimating the proportions through cluster labeling, use of contextual information resulted in better estimates when the imperfections in the labels are small. Furthermore, the use of probabilities of label imperfections resulted in better proportion estimates through cluster labeling.

9. REFERENCES

1. Anderberg, M. R.: Cluster Analysis for Applications. Academic Press (New York), 1973.

2. Everitt, B.: Cluster Analysis. John Wiley and Sons, (New York), 1974.

3. Duda, R. O.; Hart, P. E.: Pattern Classification and Scene Analysis. John Wiley and Sons (New York), 1973.

4. Heydorn, R. P.: Methods for Segment Wheat Area Estimation. Proc. LACIE Symposium. NASA/JSC, July 1979.

5. Heydorn, R. P.; Trichel, M. C.; and Erickson, J. D.: Methods for Segment Wheat Area Estimation. Proc. LACIE Symposium, vol. II, JSC 16015, July 1979, pp. 621-632.

6. Rosenfeld, A.; Hummel, R.; and Zucker, S. W.: Scene Labeling by Relaxation Operations. IEEE Trans. Systems, Man, and Cybernetics, vol. SMC-6, June 1976, pp. 420-433.

7. Rosenfeld, A.: Iterative Methods in Image Analysis. Proc. IEEE Conference on Pattern Recognition and Image Processing, June 1977, pp. 14-18.

8. Zucker, S. W.; Krishna Murty, E. V.; and Haar, R. L.: Relaxation Processes for Scene Labeling: Convergence, Speed, and Stability. IEEE Trans. Systems, Man, and Cybernetics, vol. SMC-8, Jan. 1978, pp. 41-48.

9. Chittineni, C. B.: Utilization of Spectral-Spatial Information in the Classification of Imagery Data. JSC-16335, LEMSCO 14390 (Houston), June 1980.

10. Kettig, R. L.; and Landgrebe, D. A.: Classification of Multispectral Image Data by Extraction and Classification of Homogeneous Objects. IEEE Trans. Geoscience Electronics, vol. GE-14, Jan. 1976, pp. 19-26.

11. Swain, P. H.: Bayesian Classification in a Time-Varying Environment. IEEE Trans. Systems, Man, and Cybernetics, vol. SMC-8, Dec. 1978, pp. 879-883.

12. Welch, J. R.; and Salter, K. G.: A Context Algorithm for Pattern Recognition and Image Interpretation. IEEE Trans. Systems, Man, and Cybernetics, vol. SMC-1, Jan. 1971, pp. 24-30.

13. Chittineni, C. B.: Learning With Imperfectly Labeled Patterns. Proc. Conference on Pattern Recognition and Image Processing, Aug. 1979, pp. 52-62.

44

14. Chittineni, C. B.: Maximum Likelihood Estimation of Label Imperfection Probabilities and Its Use in the Identification of Mislabeled Patterns. JSC-16067, LEC-13678, (Houston), Sept. 1979.

15. Chittineni, C. B.: Estimation of Probabilities of Label Imperfections and Correction of Mislabels. JSC-16342, LEMSCO-14356 (Houston), Mar. 1980.

16. Fletcher, R.; and Powell, M. J. D.: A Rapdily Convergent Descent Method for Minimization. Computer Journal, vol. 6, April 1963, pp. 163-168.

17. Johnson, I. L.: The Davidon-Fletcher-Powell Penalty Function Method: A Generalized Iterative Technique for Solving Parameter Optimization Problems. TND-8251, May 1976.

18. Cooper, L.; and Steinberg, D.: Introduction to Methods of Optimization. W. B. Saunders Co., 1970.

19. Lennington, R. K.; and Rassbach, M. E.: Mathematical Description and Program Documentation for Classy: An Adaptive Maximum Likelihood Clustering Method. JSC-14621, LEC-12177, (Houston), Apr. 1979.

20. Bryant, J.: On the Clustering of Multidimensional Pictorial Data. Pattern Recognition, vol. 11, no. 2, 1979, pp. 115-125.

21. Peters, B. C., Jr.: On the Consistency of the Maximum Likelihood Estimate of Normal Mixture Parameters for a Sample With Field Structure. Report no. 74, Dept. of Math., Univ. of Houston, Sept. 1979.

APPENDIX A

PROBABILITIC CLUSTER LABELING WITH FIELD STRUCTURE

APPENDIX A

PROBABILITIC CLUSTER LABELING WITH FIELD STRUCTURE

In the practical applications of pattern recognition, such as in the classification of remotely sensed agricultural imagery data, one of the difficult problems is obtaining labels for the training patterns. The labels for the training patterns are usually provided by an analyst-interpreter who examines imagery films and uses other information such as historic information and crop calendar models.

It has been observed that the field-like structures that are normally in agricultural imagery are relatively easy to label in comparison to the pixels. Recently, considerable interest has been shown in developing techniques for locating fields in the imagery data (ref. 20) and for developing maximum likelihood clustering algorithms (ref. 21) to fit the mixture of Gaussian density functions by taking the field structure of the data into account. These algorithms typically give the a priori probabilities and Gaussian cluster conditional densities for the inherent modes in the data. The situation is illustrated in the following figure.

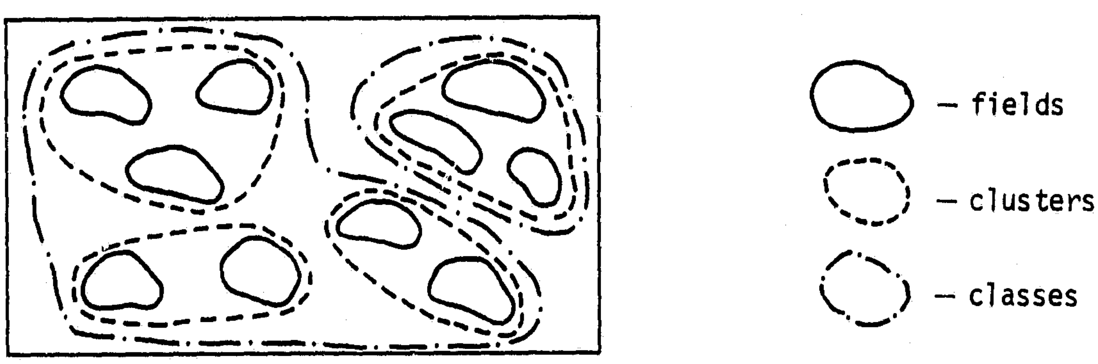

Figure A-1.- Illustration of fields, clusters, and classes in an image.

It is the purpose of this appendix to consider the problem of obtaining the probabilities of class labels for the clusters using information from a given set of labeled fields. It is assumed that a set of labeled fields from each class is given. Let $F_j(i)$, $j = 1,2,\cdots,f(i)$ be the labeled fields of class i,

$i = 1,2,\cdots,C$. Let $N_j(i)$ be the number of pixels in the j^{th} labeled field of class i. Let $x_{jk}(i)$ be the spectral vector of the k^{th} pixel of j^{th} labeled field of class i. Let $X_j(i)$ be the concatenated vector of spectral vectors of pixels in the j^{th} labeled field of class i. That is

$$X_j(i) = \begin{bmatrix} x_{j1}(i) \\ x_{j2}(i) \\ \vdots \\ x_{jN_j(i)}(i) \end{bmatrix} \tag{A-1}$$

It is also assumed that the probability density functions and a priori probabilities of the clusters are given. Let these be $p(X|\Omega = i)$ and $\delta_i, i = 1,2,\cdots,m$, respectively, where m is the number of clusters. Assuming the fields are independent, the likelihood of occurrence of $X_j(i)$ with their labels $\omega_j(i) = i$, but normalized, is given by

$$\begin{aligned} \ell &= \prod_{i=1}^{C} \prod_{j=1}^{f(i)} \frac{p[X_j(i),\omega_j(i) = i]}{p[X_j(i)]} \\ &= \prod_{i=1}^{C} \prod_{j=1}^{f(i)} p[\omega_j(i) = i|X_j(i)] \end{aligned} \tag{A-2}$$

If X is a concatenated vector of spectral vectors in a field, similar to equation (4), we have

$$p(\omega = i|X) = \sum_{\ell=1}^{m} \alpha_{\ell i} p(\Omega = \ell|X) \tag{A-3}$$

Using equation (A-3) in equation (A-2), the log likelihood function can be witten as

$$\begin{aligned} L &= \log(\ell) \\ &= \sum_{i=1}^{C} \sum_{j=1}^{f(i)} \log\left\{\sum_{\ell=1}^{m} \alpha_{\ell i} p[\Omega = \ell|X_j(i)]\right\} \end{aligned} \tag{A-4}$$

A fixed-point iteration equation for the probabilities of class labels for the clusters $\alpha_{\ell i}$ that maximize L of equation (A-4), subject to the constraints of equation (5), can be written from equations (9) and (10) as

$$\alpha_{\ell i} = \frac{\sum_{j=1}^{f(i)} d_{\ell ij}}{\sum_{r=1}^{C} \sum_{s=1}^{f(r)} d_{\ell rs}} \tag{A-5}$$

where

$$d_{\ell ij} = \frac{\alpha_{\ell i} p[\Omega = \ell | X_j(i)]}{\sum_{s=1}^{m} \alpha_{si} p[\Omega = s | X_j(i)]} \tag{A-6}$$

But from the Bayes rule, we have

$$\begin{aligned} p[\Omega = \ell | X_j(i)] &= \frac{P(\Omega = \ell) p[X_j(i) | \Omega = \ell]}{p[X_j(i)]} \\ &= \frac{P(\Omega = \ell) p[X_j(i) | \Omega = \ell]}{\sum_{s=1}^{m} P(\Omega = s) p[X_j(i) | \Omega = s]} \end{aligned} \tag{A-7}$$

The computation of a posteriori probabilities of the clusters $p[\Omega = \ell | X_j(i)]$ can be considerably simplified by noting that the sequence $[X_j(i), S_j(i),$ $j = 1,2,\cdots,f(i); i = 1,2,\cdots,C]$ is a sufficient statistic for the criterion, where $X_j(i)$ and $S_j(i)$ are the sample mean and the sample scatter matrix of the j^{th} field of the i^{th} class, respectively. That is

$$\left.\begin{aligned} \bar{X}_j(i) &= \frac{1}{N_j(i)} \sum_{k=1}^{N_j(i)} x_{jk}(i) \\ \text{and} \quad S_j(i) &= \sum_{k=1}^{N_j(i)} \left\{ [x_{jk}(i) - x_j(i)][X_{jk}(i) - X_j(i)]^T \right\} \end{aligned}\right\} \tag{A-8}$$

29

The sufficiency of the sequence $[\bar{X}_j(i), S_j(i)]$ implies that

$$\frac{P(\Omega = \ell)p[X_j(i) | \Omega = \ell]}{p[X_j(i)]} = \frac{\delta_\ell q_\ell[\bar{X}_j(i), S_j(i)]}{q[\bar{X}_j(i), S_j(i)]} \tag{A-9}$$

where $q_\ell[\bar{X}_j(i), S_j(i)]$ is the joint density of $\bar{X}_j(i)$ and $S_j(i)$, given that the cluster ℓ contains the field $F_j(i)$ and

$$q[\bar{X}_j(i), S_j(i)] = \sum_{\ell=1}^{m} \delta_\ell q_\ell[\bar{X}_j(i), S_j(i)] \tag{A-10}$$

If $p(x|\Omega = \ell) \sim N(\mu_\ell, \Sigma_\ell)$, the joint density $q_\ell[\bar{X}_j(i), S_j(i)]$ can be expressed as

$$q_\ell[\bar{X}_j(i), S_j(i)] = N_d\left[\bar{X}_j(i);\ \mu_\ell,\ \frac{1}{N_j(i)}\ \Sigma_\ell\right] W_d[S_j(i);\ N_j(i) - 1,\ \Sigma_\ell] \tag{A-11}$$

where $N_d\left(\bar{X}_j(i);\ \mu_\ell,\ \frac{1}{N_j(i)}\ \Sigma_\ell\right)$ is the d-variate normal density of $\bar{X}_j(i)$ and $W_d[S_j(i);\ N_j(i) - 1,\ \Sigma_\ell]$ is the Wishart density of $S_j(i)$ with $N_j(i) - 1$ degrees of freedom. It can easily be shown that the density of sample mean $\bar{X}_j(i)$ is given by

$$p[\bar{X}_j(i) | \Omega = \ell] \sim N\left(\mu_\ell,\ \frac{1}{N_j(i)}\ \Sigma_\ell\right) \tag{A-12}$$

The Wishart density of $S_j(i)$ with $N_j(i) - 1$ degrees of freedom can be written as

$$W_d\left[S_j(i), N_j(i) - 1, \Sigma_\ell\right] = \frac{|S_j(i)|^{(1/2)\{[N_j(i) - 1] - d - 1\}} \exp\left\{-\frac{1}{2}\ \mathrm{tr}\left[S_j(i)\ \Sigma_\ell^{-1}\right]\right\}}{2^{[1/2][N_j(i)-1]d}\ \pi^{d(d-1)/4}\ |\Sigma_\ell|^{1/2[N_j(i)-1]} \left[\prod_{u=1}^{d} \Gamma\left(\frac{1}{2}\ \{[N_j(i) - 1] + 1 - u\}\right)\right]} \tag{A-13}$$

Using equations (A-12) and (A-13) in equation (A-9) yields

$$\frac{P(\Omega = \ell)q_{\ell}[\bar{X}_j(i), S_j(i)]}{q[\bar{X}_j(i), S_j(i)]} = \frac{\delta_{\ell}\left\{|\Sigma_r|^{\frac{-N_j(i)}{2}} \exp\left[-\frac{1}{2}\,\mathrm{tr}\left(\Sigma_{\ell}^{-1}\left\{S_j(i) + N_j(i)[\bar{X}_j(i) - \mu_{\ell}][\bar{X}_j(i) - \mu_{\ell}]^T\right\}\right)\right]\right\}}{\sum_{r=1}^{m} \delta_r\left\{|\Sigma_r|^{\frac{-N_j(i)}{2}} \exp\left[-\frac{1}{2}\,\mathrm{tr}\left(\Sigma_r^{-1}\left\{S_j(i) + N_j(i)[\bar{X}_j(i) - \mu_r][\bar{X}_j(i) - \mu_r]^T\right\}\right)\right]\right\}} \tag{A-14}$$

Equation (A-14) can be used in equaitons (A-5) and (A-6) to obtain optimal probabilities of class labels for the clusters using information from a given set of labeled fields.

APPENDIX B

CONTEXTUAL CLUSTER LABELING WITH THE CRITERION OF PROBABILITY OF CORRECT LABELING

APPENDIX B

CONTEXTUAL CLUSTER LABELING WITH THE CRITERION OF PROBABILITY OF CORRECT LABELING

The problem of obtaining the optimal probabilities of class labels for the clusters using the criterion of probability of correct labeling is formulated in this appendix. It is assumed that a set of patterns $X_i(j)$ with imperfect labels $\omega_i'(j) = i$ and with the neighbors $Y_i^1(j),\cdots,Y_i^4(j)$ as shown in figure 1, for $j = 1,2,\cdots,N_i$ and $i = 1,2,\cdots,C$, are given. The probabilities of label imperfections β_{ji} are assumed to be available. It is also assumed that the probability density functions and the a priori probabilities of the clusters are given. If a pattern X with the neighbors $Y^1,\cdots,Y^4$ comes from class i, then for particular a priori probabilities and probability densities of the classes the probability with which it is correctly classified into class i is $p(\omega = i|X,Y^1,\cdots,Y^4)$. Since logorithm is a monotomic function of its argument, the criterion of probability of correct labeling (PCL) may be defined as

$$P_{CL} = \sum_{i=1}^{C} P(\omega = i) \int \log[p(\omega = i|X,Y^1,\cdots,Y^4)]p(X|\omega = i)dx \qquad \text{(B-1)}$$

Let β be the matrix of probabilities of label imperfections, where

$$\beta = [\beta_{ij}] \qquad \text{(B-2)}$$

Let

$$\nu = (\beta^T)^{-1} \qquad \text{(B-3)}$$

Using equations (B-2) and (B-3) and inverting equation (32) results in

$$P(\omega = i)p(X|\omega = i) = \sum_{j=1}^{C} \nu_{ij}\, P(\omega = j)\, p(X|\omega' = j) \qquad \text{(B-4)}$$

From equations (B-1) and (B-4) we get

$$P_{CL} = \sum_{i=1}^{C} \sum_{j=1}^{C} \nu_{ij} P(\omega' = j) \int \log[p(\omega = i|X, Y^1, \cdots, Y^4)] p(X|\omega' = j) dX \tag{B-5}$$

Using the given imperfectly labeled patterns and their neighbors, an estimate for P_{CL} of equation (B-5) can be written as

$$\hat{P}_{CL} = \sum_{i=1}^{C} \sum_{j=1}^{C} \nu_{ij} P(\omega' = j) \left(\frac{1}{N_j} \sum_{k=1}^{N_j} \log\left\{ p\left[\omega = i | X_j(k), Y_j^1(k), \cdots, Y_j^4(k) \right] \right\} \right) \tag{B-6}$$

Substituting sample estimates for $P(\omega' = j)$ and using equations (24) and (4) in equation (B-6), the criterion can be written as

$$\begin{aligned} Cr = & \sum_{i=1}^{C} \sum_{j=1}^{C} \nu_{ij} \sum_{k=1}^{N_j} \log\left\{ \sum_{r=1}^{m} \alpha_{ri} p[\Omega = r | X_j(k)] \right\} \\ & + \sum_{u=1}^{C} \sum_{j=1}^{C} \nu_{uj} \sum_{k=1}^{C} \sum_{\ell=1}^{4} \log\left\{ \sum_{k_\ell=1}^{C} \sum_{r=1}^{m} \alpha_{rk_\ell} \xi_{k_\ell u} p[\Omega = r | Y_j^\ell(k)] \right\} \end{aligned} \tag{B-7}$$

where

$$\xi_{k_\ell u} = \frac{P\left[\omega_j^\ell(k) = k_\ell | \omega_j(k) = u \right]}{P\left[\omega_j^\ell(k) = k_\ell \right]} \tag{B-8}$$

The probabilities α_{ri} that maximize Cr of equation (B-7) and that are subject to the constraints of equation (5) can be obtained using optimization techniques such as Davidon-Fletcher-Powell (refs. 16-18). However, fixed-point iteration equations are developed in the following.

B.1 FIXED-POINT ITERATION SCHEME FOR OPTIMAL α_{ri}

It is noted that in equation (B-7), ν_{ij} might be negative. In the following fixed-point iteration equations for obtaining optimal α_{ri}, the probabilities of class labels for the clusters are developed. Consider

$$p(X|\omega = i) = \frac{1}{P(\omega = i)} \sum_{j=1}^{C} p(X,\omega = i,\omega' = j)$$

$$= \sum_{j=1}^{C} \beta_{ij}\, p(X|\omega' = j) \tag{B-9}$$

where it is assumed that

$$p(X|\omega = i,\ \omega' = j) = p(X|\omega' = j) \tag{B-10}$$

In terms of probabilities of label imperfections, the a priori probabilities of perfect and imperfectly labeled classes are related as

$$P(\omega' = i) = \sum_{j=1}^{C} \beta_{ji} P(\omega = j) \tag{B-11}$$

Inverting equation (B-11), we get

$$P(\omega = i) = \sum_{j=1}^{C} \nu_{ij} P(\omega' = j) \tag{B-12}$$

Using equation (B-9) in equation (B-1), an estimate for P_{CL} can be written as

$$\hat{P}_{CL} = \sum_{i=1}^{C} \sum_{j=1}^{C} \eta_{ij} \sum_{k=1}^{N_j} \left\{ p\left[\omega = i|X_j(k), Y_j^1(k), \cdots, Y_j^4(k)\right]\right\} \tag{B-13}$$

where

$$\eta_{ij} = \frac{P(\omega = i)\beta_{ij}}{N_j} \tag{B-14}$$

From equations (4), (24), and (B-13), the criterion becomes

$$Cr = \sum_{i=1}^{C} \sum_{j=1}^{C} n_{ij} \sum_{k=1}^{N_j} \log\left\{ \sum_{r=1}^{m} \alpha_{ri} p[\Omega = r|X_j(k)] \right\}$$

$$+ \sum_{u=1}^{C} \sum_{j=1}^{C} n_{uj} \sum_{k=1}^{N_j} \sum_{\ell=1}^{4} \log\left\{ \sum_{k_\ell=1}^{C} \sum_{r=1}^{m} \alpha_{rk_\ell} \xi_{k_\ell u} p\left[\Omega = r|Y_j^\ell(k)\right] \right\} \tag{B-15}$$

The following fixed-point iteration equations for obtaining optimal α_{ri} that maximize Cr of equation (B-15), subject to the constraints of equation (5), can easily be obtained by introducing Lagrangian multipliers. That is

$$\alpha_{ri} = \frac{\alpha_{ri}\left(\delta_{ri}^1 + \delta_{ri}^2\right)}{\sum_{i=1}^{C} \alpha_{ri}\left(\delta_{ri}^1 + \delta_{ri}^2\right)} \tag{B-16}$$

where

$$\delta_{ri}^1 = \sum_{j=1}^{C} n_{ij} \sum_{k=1}^{N_j} \frac{p[\Omega = r|X_j(k)]}{\sum_{s=1}^{m} \alpha_{si}\, p[\Omega = s|X_j(k)]} \tag{B-17}$$

and $$\delta_{ri}^2 = \sum_{u=1}^{C} \sum_{j=1}^{C} n_{uj} \sum_{k=1}^{N_j} \sum_{\ell=1}^{4} \left\{ \frac{\xi_{iu} p\left[\Omega = r|Y_j^\ell(k)\right]}{\sum_{k_\ell=1}^{C} \sum_{s=1}^{m} \alpha_{sk_\ell} \xi_{k_\ell u} p\left[\Omega = s|Y_j^\ell(k)\right]} \right\} \tag{B-18}$$

If the spatial information is not used, the fixed-point iteration equations for obtaining the optimal probabilities α_{ri} become the following.

$$\alpha_{ri} = \frac{\alpha_{ri}\, \delta_{ri}^1}{\sum_{i=1}^{C} \alpha_{ri}\, \delta^1{}_{ri}} \tag{B-19}$$

216

Where δ^1_{ri} is given by equaion (B-17). It is noted that when there are no imperfections in the labels, equation (B-19) is identical to equation (9).

B.2 EXPERIMENTAL RESULTS

This section presents some results from the processing of remotely sensed multispectral scanner imagery data. The objective of the processing is to estimate the proportion of class of interest through probabilistic cluster labeling. The class of interest is wheat and its proportion is estimated using equation (17). The same labeled patterns and the cluster statistics of section 7 are used. The a priori probabilities of imperfectly labeled classes for use in equation (B-12) are estimated as sample estimates. The a priori and the transition probabilities used in the local neighborhood of the given labeled patterns are given in equation (49). The results are listed in table B-1. From table B-1, it is seen that better proportion estimates are obtained by taking the imperfections in the labels into account.

TABLE B-1.- ESTIMATED PROPORTION OF CLASS 1 WITH IMPERFECT LABELS AND β-MATRIX

Segment	Location (county, state)	Computed β-matrix comparing A.I. and G.T. labels	Proportion estimate using eqs. (B-16), (B-17) and (B-18)	Proportion estimate using eqs. (B-19) and (B-17)	Proportion estimate directly with AI labels using eqs. (9)	G.T. proportion
1005	Sherman, Texas	$\begin{bmatrix} 0.5455 & 0.4545 \\ 0.0308 & 0.9692 \end{bmatrix}$	0.3194	0.3641	0.2456	0.348
1060	Cheyenne, Colorado	$\begin{bmatrix} 0.5667 & 0.4343 \\ 0.0263 & 0.9737 \end{bmatrix}$	0.2297	0.2787	0.1975	0.231
1231	Jackson, Oklahoma	$\begin{bmatrix} 0.9315 & 0.0685 \\ 0.1304 & 0.8696 \end{bmatrix}$	0.7640	0.7546	0.6265	0.744
1520	Big Stone, Montana	$\begin{bmatrix} 0.7917 & 0.2083 \\ 0.0149 & 0.9851 \end{bmatrix}$	0.2398	0.2661	0.2109	0.301
1604	Renville, North Dakota	$\begin{bmatrix} 0.4600 & 0.5400 \\ 0.1569 & 0.8431 \end{bmatrix}$	0.4981	0.5035	0.2963	0.524
1675	McPherson, South Dakota	$\begin{bmatrix} 0.2667 & 0.7333 \\ 0.0390 & 0.9610 \end{bmatrix}$	0.2681	0.2448	0.1085	0.291
1805	Gregory, South Dakota	$\begin{bmatrix} 0.4211 & 0.5789 \\ 0.0640 & 0.9360 \end{bmatrix}$	0.1502	0.1385	0.1181	0.164
1853	Ness, Kansas	$\begin{bmatrix} 0.8077 & 0.1923 \\ 0.0615 & 0.9385 \end{bmatrix}$	0.2769	0.3164	0.3246	0.306
Bias			0.20350E-01	0.52875E-02	0.9763E-01	
Mean square error			0.89969E-03	0.89725E-03	0.1514E-01	

NASA-JSC

38